CompCytogen 10(1): 1–15 (2016)
doi: 10.3897/CompCytogen.v10i1.5304
http://compcytogen.pensoft.net

RESEARCH ARTICLE

COMPARATIVE Cytogenetics
A peer-reviewed open-access journal
International Journal of Plant & Animal Cytogenetics, Karyosystematics, and Molecular Systematics

Screening and chromosome localization of two cotton BAC clones

Xinglei Cui[1], Fang Liu[1], Yuling Liu[1], Zhongli Zhou[1], Chunying Wang[1], Yanyan Zhao[1], Fei Meng[1], Xingxing Wang[1], Xiaoyan Cai[1], Yuhong Wang[1], Renhai Peng[1,2], Kunbo Wang[1]

1 *State Key Laboratory of Cotton Biology (China)/Institute of Cotton Research of Chinese Academy of Agricultural Science, Anyang, Henan, 455000, China* **2** *Anyang Institute of Technology, Anyang, Henan, 455000, China*

Corresponding authors: *Renhai Peng* (aydxprh@163.com); *Kunbo Wang* (wkbcri@163.com)

Academic editor: *A. Joachimiak* | Received 14 May 2015 | Accepted 30 September 2015 | Published 22 January 2016

http://zoobank.org/000A4CED-045B-47BF-80AA-3AAE4733D655

Citation: Cui X, Liu F, Liu Y, Zhou Z, Wang C, Zhao Y, Meng F, Wang X, Cai X, Wang Y, Peng R, Wang K (2016) Screening and chromosome localization of two cotton BAC clones. Comparative Cytogenetics 10(1): 1–15. doi: 10.3897/CompCytogen.v10i1.5304

Abstract

Two bacterial artificial chromosome (BAC) clones (350B21 and 299N22) of Pima 90-53 cotton [*Gossypium barbadense* Linnaeus, 1753 (2n=4x=52)] were screened from a BAC library using SSR markers. Strong hybridization signals were detected at terminal regions of all A genome (sub-genome) chromosomes, but were almost absent in D genome (sub-genome) chromosomes with BAC clone 350B21 as the probe. The results indicate that specific sequences, which only exist at the terminal parts of A genome (sub-genome) chromosomes with a huge repeat number, may be contained in BAC clone 350B21. When utilizing FISH with the BAC clone 299N22 as probe, a pair of obvious signals was detected on chromosome 13 of D genome (sub-genome), while strong dispersed signals were detected on all A genome (sub-genome) chromosomes. The results showed that peculiar repetitive sequence, which was distributed throughout all A genome (sub-genome) chromosomes, may exist in BAC clone 299N22. The absence of the repetitive sequences, which exist in the two BAC clones, in D genome may account for the genome-size variation between A and D genomes. In addition, the microcolinearity analysis of the clone 299N22 and its homologous region on *G. raimondii* Ulbrich, 1932 chromosome 13 ($D_5$13) indicated that the clone 299N22 might come from A sub-genome of sea island cotton (*G. barbadense*), and a huge number of small deletions, illegitimate recombination, translocation and rearrangements may have occurred during the genus evolution. The two BAC clones studied here can be used as cytological markers but will be also be helpful to research in cotton genome evolution and comparative genomics.

Copyright *Xinglei Cui et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords
Cotton, BAC, FISH, cytological marker, microcolinearity

Introduction

Cotton (*Gossypium* Linnaeus, 1753) provides an excellent model system for studies on polyploidization, genomic organization, and genome-size variation (Wang et al. 2010). The genus of cotton is known to be cultivated in over 100 countries and has been classified into eight diploid (2n=2x=26) genomic groups: A, B, C, D, E, F, G, K, and one allotetraploid (2n=4x=52) genomic group: AD (Percival et al. 1999). Approximately 5 MYA (million years ago) A and D genome diploids diverged, then later became reunited with allopolyploid formation 1–2 MYA (Cronn et al. 2002; Senchina et al. 2003). The latest research shows that the genome size of an A genome species is larger than that of a D genome species (Wang et al. 2012; Li et al. 2014). Many influential factors, such as polyploidization (Wendel 2000), transposable element amplification (Bennetzen 2002; Kidwell et al. 2002; Piegu et al. 2006), tandem repeat expansion (Ellegren et al. 2002; Morgante et al. 2002), gene duplication (Zhang 2003), organellar transfer to the nucleus (Shahmuradov et al. 2003), and intron size expansion (Deutsch et al. 1999; Vinogradov et al. 1999) are thought to be collectively responsible for the genome-size variation (Grover et al. 2007). Accumulation of different transposable elements classes among different genomes was thought to be the most important reason (Hawkins et al. 2006). The studies on genome-size differences between A and D genomes will help in understanding cotton evolution as well as facilitating genetic improvement of cotton.

The introduction of fluorescence *in situ* hybridization (FISH), involving hybridization of labeled DNA probes to cytological targets, such as metaphase chromosomes, interphase nuclei, and extended DNA fibers, marked the beginning of a new era for studies on chromosome structure and function. Modern methodologies and modifications, such as the development of probes from specificity for highly repeated sequences to single-copy sequence (Desel et al. 2001; Zhu et al. 1999), and from single-colored probes to multiple-colored probes (Tang et al. 2009), have all been designed to optimize the probe detection sensitivity. Nowadays, FISH is a versatile and accurate tool for chromosome localization of sequences (Gomez et al. 1997), cytogenetic map construction (Sun et al. 2013; Han et al. 2011; Cui et al. 2015), genome structure study (Zhao et al. 2011; Wang et al. 2001a), genome evolution (Wu et al. 2013), and comparative genomics study (Gan et al. 2013).

Eukaryotic genomes, with rare exceptions, are replete with interspersed repetitive DNAs, of which most are transposable elements (Feschotte 2008). Large-scale DNA sequencing has revealed that genome size is highly correlated with transposable element content (Oliver et al. 2013). The genomes of *G. arboretum* Linnaeus, 1753 and *G. raimondii* Ulbrich, 1932 have been sequenced and assembled, the comparison

between the two genomes showed the transposable elements, especially LTR, activities substantially contributed to the twofold genome-size variation (Wang et al. 2012; Li et al. 2014). In this study, two BAC clones with genome-specific repetitive sequences (350B21 and 299N22) were localized and microcolinearity of BAC clone 299N22 and its homologous region on chromosome $D_5$13 was analyzed.

Material and methods

Materials

The plant materials were obtained from National Wild Cotton Nursery in Hainan Island, China, sponsored by the Institute of Cotton Research of Chinese Academy of Agricultural Sciences (CRI-CAAS). They are also conserved in the greenhouse at CRI-CAAS' headquarter in Anyang City, Henan Province, China.

Chromosome-specific BAC clones (Wang et al. 2007) used to identify the individual chromosomes were kindly provided by Prof. Tianzhen Zhang (Nanjing Agricultural University, China).

The *G. raimondii* genome sequence was downloaded from the sequenced genome of land plants in Phytozome (http://www.phytozome.net). The *G. arboreum* genome sequence was downloaded from Cotton Genome Project (CGP: http://cgp.genomics.org.cn).

Screening of BAC library

Pima 90–53 (*G. barbadense*) BAC library screened in this paper was kindly provided by Prof. Zhiying Ma (Hebei Agricultural University, China). The simple sequence repeat (SSR) markers were selected from 3 genetic maps (Table 1) (Nguyen et al. 2004; Zhao et al. 2012; Han et al. 2006) and used to screen the BAC library. To facilitate PCR screening, a rapid method of screening BAC libraries was used to obtain positive BAC clones (Cheng et al. 2012). First, one-dimensional pools (plate pools)

Table 1. SSR markers and their genetic maps.

SSR marker	Genetic map of cotton
NAU1215	Han et al. (2006) Theor Appl Genet
CIR342	Han et al. (2006) Theor Appl Genet
NAU1023	Han et al. (2006) Theor Appl Genet
NAU1201	Han et al. (2006) Theor Appl Genet
NAU3022	Zhao et al. (2012) BMC Gnomics
NAU3384	Zhao et al. (2012) BMC Gnomics
NAU5100	Zhao et al. (2012) BMC Gnomics
CIR096	Nguyen et al. (2004) Theor Appl Genet

were made; 384 clones were pooled together on a same plate. Then, bacterial colony PCR was used to screen one-dimensional pools. Secondly, two-dimensional pools (line pools) were made and used to screen, in each of which 24 clones in a same line were pooled together. Thirdly, each clone was screened for the target DNA. Bacterial colony PCR was carried out with 1μL of Bacterial colony template in the presence of 0.5 μL of dNTPs (10mM), 0.5 U Taq DNA polymerase, 1.0 μL 10×Reaction buffer and 0.5 μL of each primer, for a final volume of 10 μL. Following initial denaturation at 95 °C for 3 min, 30 cycles of 94 °C for 45 s, annealing temperature for 45 s and 72 °C for 1 min was performed. PCR products were separated by 0.8% polyacrylamide gel electrophoresis.

DNA probe preparation

The BAC clone DNA was isolated using a standard alkaline extraction (Sambrook et al. 2002). The chromosome-specific BAC clones were labeled with digoxigenin-dUTP via nick translation, whereas the screened BAC clones were labeled with biotin-dUTP via nick translation, according to the instructions of the manufacturer (Roche Diagnostics, USA).

Chromosome preparation and FISH

Mitotic chromosome preparation and FISH procedures were conducted using a modified protocol (Wang et al. 2001b). Biotin-labeled and digoxigenin-labeled probes were detected by avidin-fluorescein (green) and anti-digoxigenin-rhodamine (red) (Roche Diagnostics, USA), respectively. Chromosomes were counterstained by 4',6-diamidino-2-phenylindole (DAPI) in antifade VECTASHIELD solutions (Vector Laboratories, Burlingame, CA). The concentration of block DNA (genomic DNA) was 200 times that of the chromosome-specific BAC DNA. The hybridization signals were observed using a fluorescence microscope (Leica MRA2) with a charge-coupled device (CCD) camera. Final image adjustments were performed using Adobe Photoshop CS3 software.

BAC clone sequencing and microcolinearity analysis

Both BAC clone 350B21 and 299N22 were outsourced to a biological company for sequencing. The sequences of BAC clones were used as query sequences to search for its homologous regions using BLASTN algorithms against A_2 genome and D_5 genome. Microcolinearity analysis of homologous regions was achieved using software CIRCOS.

Results

Identification and selection of cotton BAC clones

A total of 192 plate pools (73728 BAC clones, nearly covering *G. barbadense* genome 3 times) were constructed and screened using bacterial colony PCR. Nineteen positive BAC clones were identified (Table 2) and selected to be probes for FISH. Seventeen clones, which showed ambiguous FISH signals or no FISH signal on *G. barbadense* mitotic metaphase chromosomes, were discarded. BAC clones 350B21 and 299N22, which showed obvious characteristic signals on *G. barbadense* mitotic metaphase chromosomes, were selected for further study.

Localization of BAC 350B21

Obvious signals were detected on terminal parts of all *G. barbadense* Linnaeus, 1753 (A_2D_2, 2n=4x=52) A sub-genome chromosomes with BAC clone 350B21 as probe. And signals were alike when using four other tetraploid species [*G. hirsutum* Linnaeus, 1753 (A_1D_1, 2n=4x=52), *G. tomentosum* Nuttall ex Seemann, 1865 (A_3D_3, 2n=4x=52), *G. mustelinum* Miers ex Watt, 1907 (A_4D_4, 2n=4x=52), *G. darwinii* Watt, 1907 (A_5D_5, 2n=4x=52)] mitotic metaphase chromosomes as target DNAs. Then, mitotic metaphase chromosomes of two A genome species [*G. arboretum* (A_1, 2n=2x=26), *G. herbaceum* Linnaeus, 1753 (A_2, 2n=2x=26)] were used as target DNAs and obvious signals were detected at terminal parts of all the chromosomes. On the contrast, no obvious signal, except two pair of weak signals, was detected on chromosomes of two D genome cotton species [*G. thurberi* Todaro, 1878 (D_1, 2n=2x=26) and *G. raimondii* (D_5, 2n=2x=26)]. The signals were alike between A genomes and A sub-genomes as well as D genomes and D sub-genomes (Fig. 1).

Table 2. Screened clones of Pima 90-53 BAC library.

SSR markers	Screened clones from BAC library
NAU1215	300N10
CIR342	268E2; 268K2
NAU1023	311A4; 311A11
NAU1201	299N22; 323O3; 317K24; 185N14
NAU3022	30A18; 106P24
NAU3384	328L13
NAU5100	389J15; 376M12; 311M1
CIR096	399A22; 162G3; 350B21;342O11

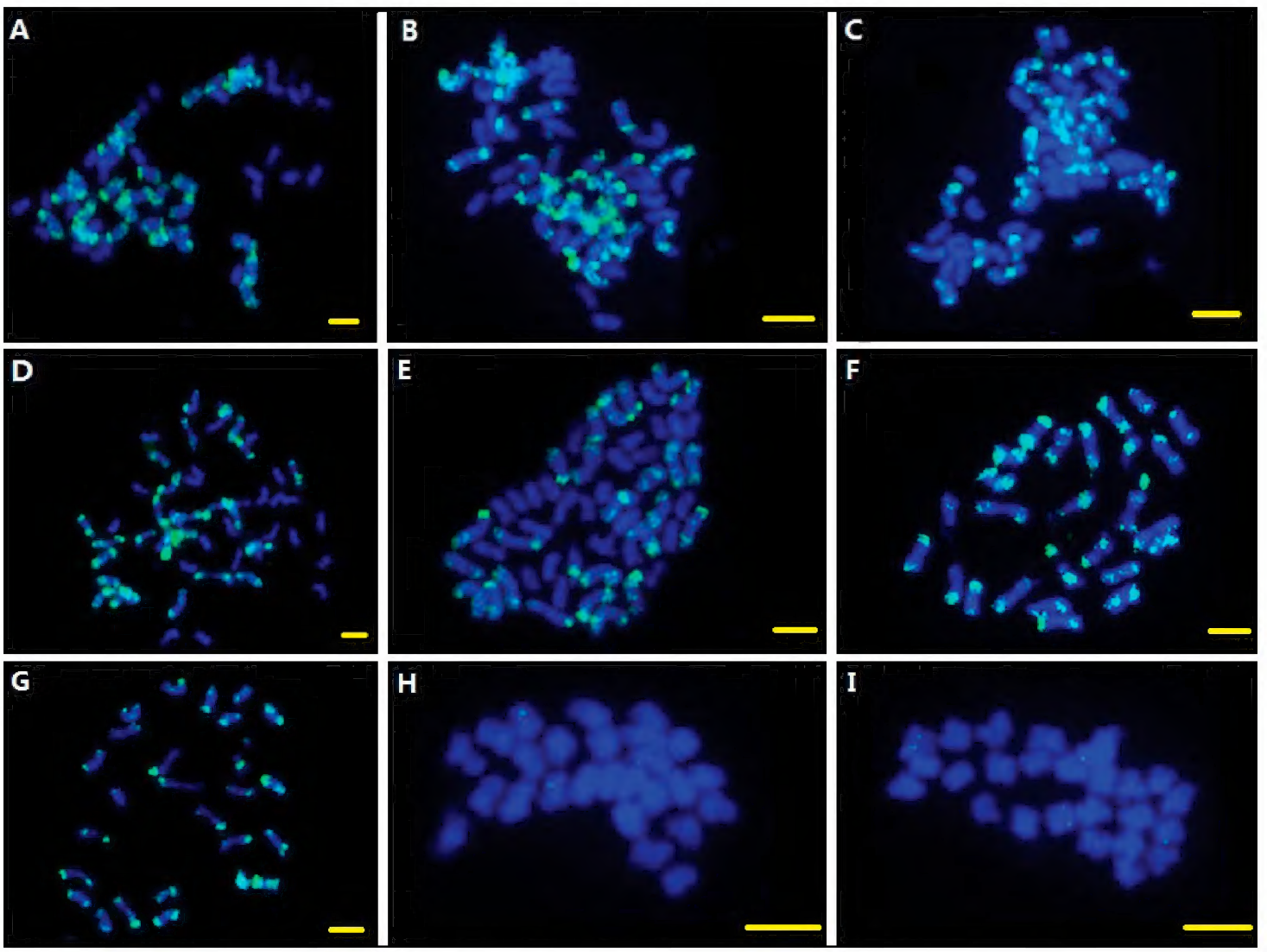

Figure 1. The FISH images of BAC clone 350B21(green) hybridized to mid-mitotic chromosomes in different *Gossypium* species, Bar: 5 μm. **A** *G. hirsutum* (A_1D_1, 2n=4x=52) **B** *G. barbadense* (A_2D_2, 2n=4x=52); **C** *G. tomentosum* (A_3D_3, 2n=4x=52) **D** *G. mustelinum* (A_4D_4, 2n=4x=52) **E** *G. darwinii* (A_5D_5, 2n=4x=52) **F** *G. arboretum* (A_1, 2n=2x=26) **G** *G. herbaceum* (A_2, 2n=2x=26) **H** *G. thurberi* (D_1, 2n=2x=26) **I** *G. raimondii* (D_5, 2n=2x=26).

Localization of BAC 299N22

Obvious disperse signals were detected on all A sub-genome chromosomes of tetraploid species [*G. hirsutum* (A_1D_1, 2n=4x=52), *G. barbadense* (A_2D_2, 2n=4x=52), *G. tomentosum* (A_3D_3, 2n=4x=52), *G. mustelinum* (A_4D_4, 2n=4x=52), *G. darwinii* (A_5D_5, 2n=4x=52)] with BAC clone 299N22 as probe. When mitotic metaphase chromosomes of two A genome species [*G. arboretum* (A_1, 2n=2x=26), *G. herbaceum* (A_2, 2n=2x=26)] were used as target DNAs, obvious signals were detected on all the chromosomes, while only a pair of obvious signals was detected on chromosome 13 of two D genome cotton species [*G. thurberi* (D_1, 2n=2x=26) and *G. raimondii* (D_5, 2n=2x=26)]. The relative position of FISH signals on chromosome $D_5$13 was measured to be about 62.4FL (FL: the percentage of the distance from the FISH site to the end of the short arm relative to the total length of the chromosome) after measuring more than 10 cells with clear chromosome spreads (Fig. 2).

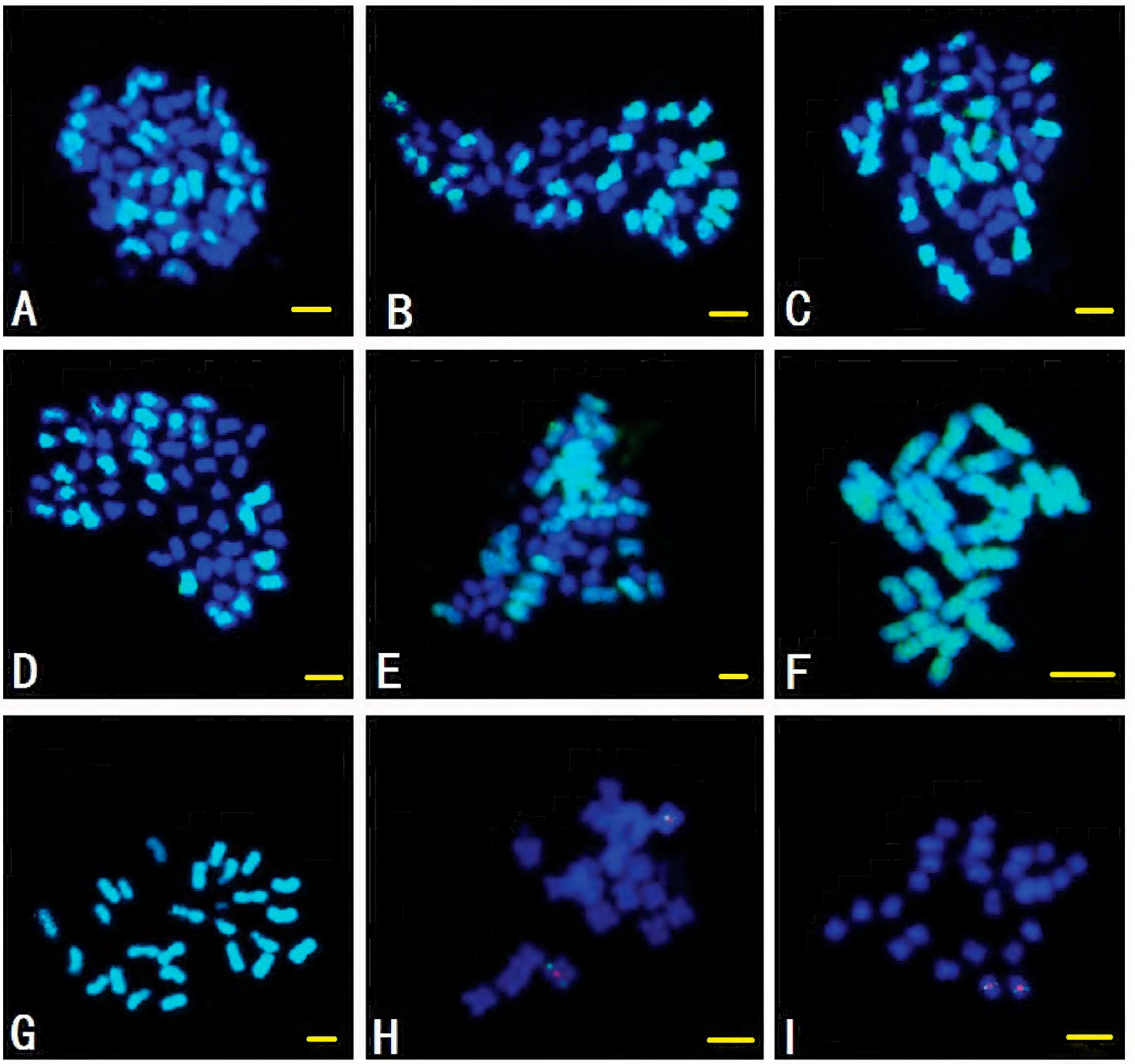

Figure 2. The FISH images of BAC clone 299N22 (green) hybridized to mid-mitotic chromosomes different *Gossypium* species, Bar: 5 μm. **A** *G. hirsutum* (A_1D_1, 2n=4x=52) **B** *G. barbadense* (A_2D_2, 2n=4x=52) **C** *G. tomentosum* (A_3D_3, 2n=4x=52) **D** *G. mustelinum* (A_4D_4, 2n=4x=52) **E** *G. darwinii* (A_5D_5, 2n=4x=52) **F** *G. arboretum* (A_1, 2n=2x=26); **G** *G. herbaceum* (A_2, 2n=2x=26) **H** *G. thurberi* (D_1, 2n=2x=26) **I** *G. raimondii* (D_5, 2n=2x=26). Red: the signal of chromosome-specific BAC clone for chromosome $D_1$13, $D_5$13.

Analysis of BAC clones sequences and microcolinearity

Sequencing of BAC clone 350B21 failed, as too many simple repeat sequences existed in the BAC clone. A new lineage-specific LTR family, which accounted for about 35% of A_2 genome while being absent in D_5 genome, was identified analyzing the sequence of BAC 299N22. The sequence of BAC clone 299N22 was used as query sequence to search for its homologous regions using BLASTN algorithms against A_2 genome (*G. arboretum*) and D_5 (*G. raimondii*) genome, respectively. When A_2 genome was used as a database, multiple dispersedly distributed hits on all chromosomes of A_2 genome were obtained (Fig. 3A), so the homologous region of BAC 299N22 in A_2 genome was

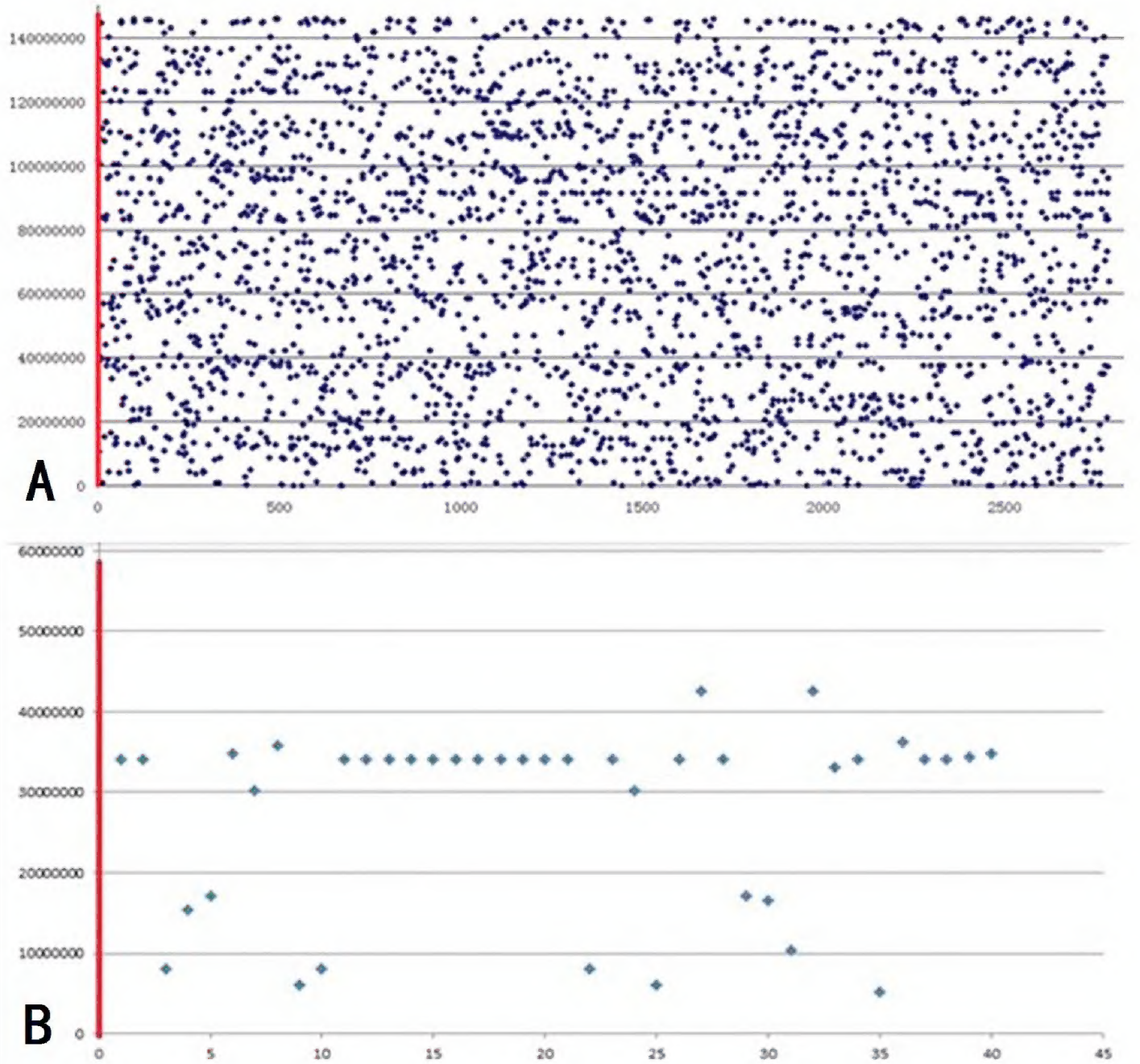

Figure 3. The distribution of BAC 299N22 clone on chromosomes. X-coordinate indicates the length of chromosome, y-coordinate indicates the hits of sequence alignment of BAC clone 299N22 and chromosomes. **A** the result of BLASN, with chromosome $A_2$02 as database **B** the result of BLASTN, with chromosome $D_5$13 as database.

not identified. When D_5 genome was used as a database, similar sequences were only detected in chromosome 13, and the density was obviously higher at the region of 34067000bp—34098000bp (58.41% of chromosome $D_5$13, the position was almost the same as FISH result) than that of other regions of $D_5$13 chromosome. When the E value was set lower, the hits were only found in that region (Fig. 3B). Therefore, the 31kb region on chromosome $D_5$13 was thought to be the homologous region of BAC 299N22. Using the CIRCOS software analysis, the microcolinearity of BAC clone 299N22 and its homologous region on chromosome $D_5$13 proved to be poor. The orders of the highly conserved fragments showed discrepancies, even the orientations of some highly conserved fragments were different. According to cotton SSR primer sequence information on NCBI and *G. raimondii* genome annotation information, 4 SSR markers, NAU1201, NAU1141, HAU3220, and MON_CGR5697, and 2 genes

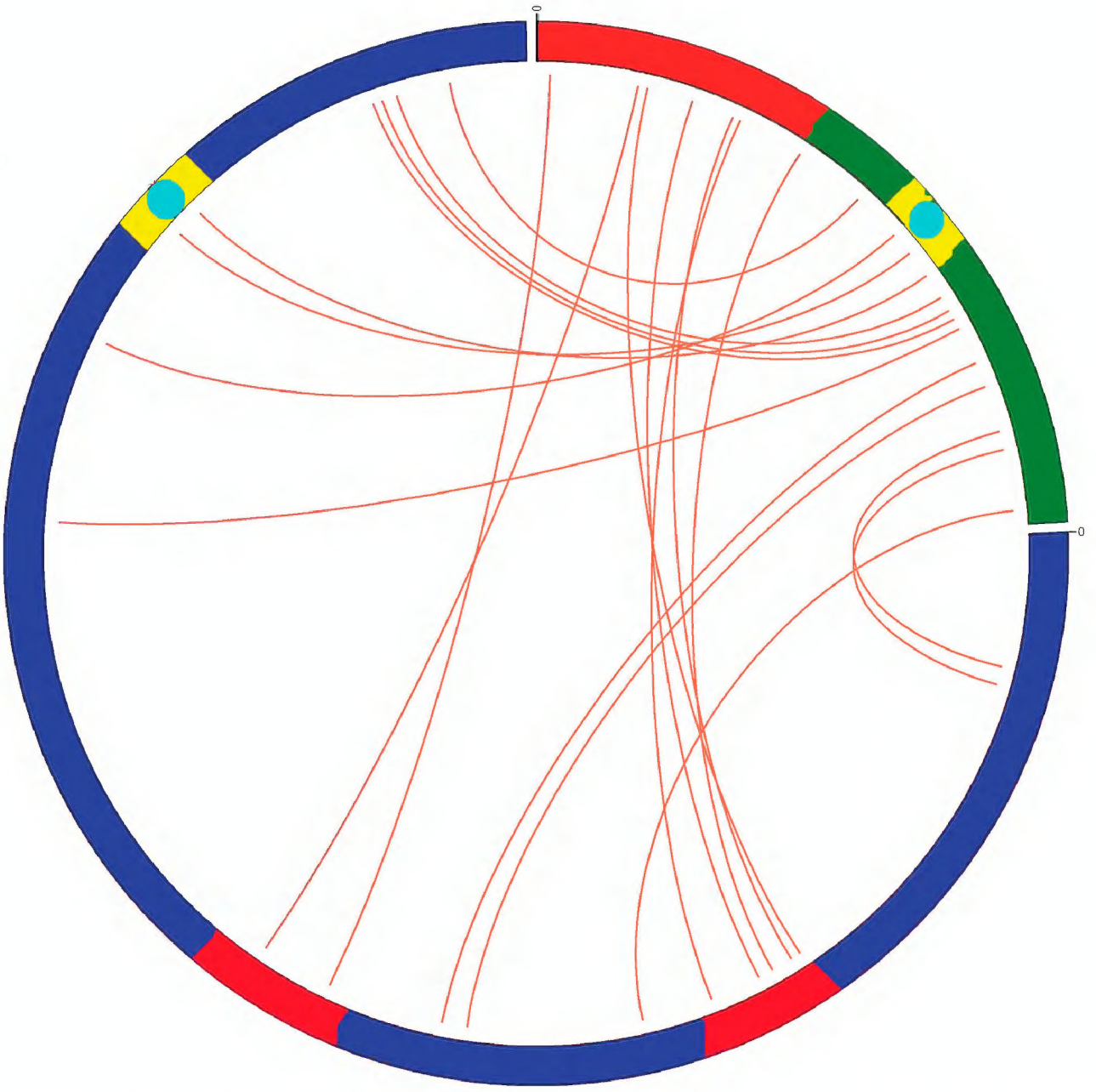

Figure 4. Microcolinearity of the BAC 299N22 and its homologous fragment on chromosome $D_5$13 of *G. raimondii*. Lower-left is BAC 299N22, top-right is its homologous fragment on chromosome $D_5$13 of *G. raimondii*. Red: gene 013G130100, yellow: 013G130200, light blue: 4 SSR markers.

013G130100 and 013G130200, were contained in the pair of homologous regions (Fig. 4). However, the distance between the two genes in BAC clone 299N22 was much longer than that on $D_5$13, and gene 013G130100 in BAC clone 299N22 was divided into two segments. The results also showed special sequences, which do not exist in D genome, repeated a huge number of times in A genome, exist in BAC 299N22.

Discussion

New cytological markers

Chromosome identification is the foundation of research on plant genetics, evolution and genomics. Conventional individual chromosome identification is mainly based on

analyzing chromosomal relative lengths and arm ratios, and, as a result, is very difficult and inaccurate when identifying chromosomes small and similar. Therefore finding suitable molecular cytogenetic markers becomes very necessary for the unambiguous identification of individual chromosomes. FISH is a reliable cytological technique for chromosome identification, and has been adapted successfully to identify the chromosomes for many plant species, including rice (Cheng et al. 2001), potato (Dong et al. 2000), sorghum (Kim et al. 2005) and so on. A set of chromosome-specific BAC clones for *G. hirsutum* chromosomes identification has been developed and was applied successfully in many cotton species (Wang et al. 2007; Wang et al. 2008; Gan et al. 2011; Gan et al. 2012). In this study, BAC clone 299N22 could be a new cytological marker for chromosome 13 of D genome (sub-genome), and its cytogenetic position was measured to be approximately 62.4 FL. As BAC clone 299N22 showed well-distributed repetitive signals on all A genome (sub-genome) chromosomes, it also could be used as a cytological marker for identifying A genome (sub-genome) chromosomes. BAC clone 350B21, which showed repetitive signals at the terminal regions of all A genome (sub-genome) chromosomes could be used as a cytological marker for identifying or labeling terminal regions of all A genome (sub-genome) chromosomes. The addition of these new cytological markers will facilitate the study of cotton genomics and evolution.

Cotton A genome (sub-genome) has unique repetitive sequences

Repetitive DNA sequences form a large portion of the genomes of eukaryotes, indicating a major contributor to variation in genome size among organisms of similar complexity (Charlesworth et al. 1994). The genus *Gossypium*, which provides a facile system for investigating the genomic organization and evolution, also has a high content of repetitive sequences in its genome. Different types of repeat sequences accounted for as much as 68.5% of the *G. arboreum* genome and approximately 57% of the *G. raimondii* genome, respectively. And most of the repetitive sequences are long terminal repeat (LTR) retrotransposons (Wang et al. 2012; Li et al. 2014).

When using BAC clone 350B21 as a probe, strong signals were detected at the terminal parts of all chromosomes of A genome (sub-genome), while being absent on D genome (sub-genome) chromosomes. The results may indicate that special repetitive sequences in BAC clone 350B21 have a bias of insertion sites at terminal parts of A genome (sub-genome) chromosomes. Another kind of repetitive sequence exists in BAC clone 299N22 showed well-distributed dispersed signals on all A genome (sub-genome) chromosomes. These unique repetitive sequences may be the major reason for the genome-size difference between A genome and D genome.

A new LTR family, which accounts for about 35% of A_2 genome while almost being absent in D_5 genome, was identified analyzing the sequence of BAC clone 299N22. The LTR family was inserted randomly along each chromosome in *G. arboretum* genome, and was different from any reported repetitive sequences in cotton

(Hawkins et al. 2006). As the LTR family accounts for so much of A genome, it should be different from any sequence reported by Zhao et al (1995) and Hanson et al (1998). The identification of the new LTR family will facilitate understanding of the differences between the two genomes. Sequencing of BAC clone 350B21 failed as too many simple repeat sequences existed in the BAC clone. This indicates that the terminal regions of A genome (sub-genome) chromosomes may be replete with simple repeat sequences. Their absence in D genome (sub-genome) indicates that they may appear after the divergence of A, D genomes and contributed to the genome-size difference between the two genomes. The similarity of signals in D genome and D sub-genome suggests that the repetitive sequences in the two BAC clones may not occur colonization after polyploidization event, this indicate they may turned to be silent before the polyploidization event.

Many factors contributed to genome-size evolution.

Many factors are thought to be responsible for the genome-size variation. The analysis of *AdhA* and *CesA* regions of different cotton genomes indicated that many forces operated collectively among genomic regions to reflect genome-size evolution (Grover et al. 2007). The microcolinearity analysis, comparative analysis of homologous sequences from different genomes, is a method of comparative genomics research for studying and speculating upon the relationships between genomes and evolution patterns. In the present study, the homologous region of BAC clone 299N22 on chromosome $D_5$13 was obtained using bioinformatics analysis. As the sequence of BAC clone 299N22 is much longer than its homologous region on chromosome $D_5$13, BAC clone 299N22 was thought to be from A sub-genome of *G. barbadense*. Microcolinearity analysis of the homologous regions showed that the orders of the most highly conserved fragments were different, even the orientations of some highly conserved fragments was different, which may indicate that a large number of translocations, inversions, and segmental rearrangements occurred during evolution. The analysis showed the length of gene parts appeared similar between the homologous regions, while gene-free regions were not. This may provide a hint that the evolution between gene islands or in gene-free regions may be the main reason for the genome-size variations, as previously reported (Grover 2004). The repetitive sequences which were distributed dispersedly on A genome chromosomes were located at the non-genetic regions, and this may indicate that the difference in non-genetic regions may be attributed to the accumulation of repetitive sequences.

Conclusions

In recent years, many achievements, such as in the study of cytogenetic map construction, genome evolution, and comparative genomics, have been obtained by using

BAC-FISH. The repetitive sequences in the two BAC clones showed distribution bias and may be an important reason for the genome-size variation. Analysis of the repetitive sequences will be helpful in the studies on cotton genome evolution and comparative genomics.

Acknowledgements

We deeply thank Prof. Tianzhen Zhang (Nanjing Agricultural University, China) for providing the set of chromosome-specific BAC clones, Prof. Zhiying Ma (Heibei Agricultural University, China) for supplying the BAC library. The research was sponsored by a grant from the National Natural Science Foundation of China (No. 31471548), State Key Laboratory of Cotton Biology Open Fund (No.CB2014A07), National High Technology Research and Development Program (No.2013AA102601).

References

Bennetzen JL (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 115: 29–36. doi: 10.1023/A:1016015913350

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215–220. doi: 10.1038/371215a0

Cheng H, Peng RH, Zhang XD, Liu F, Wang CY, Wang KB (2012) A rapid method to screen BAC library in cotton. Biotechnology 22: 55–57.

Cheng ZK, Buell CR, Wing RA, Gu MH, Jiang JM (2001) Toward a cytological characterization of the rice genome. Genome Research 11: 2133–2141. doi: 10.1101/gr.194601

Cronn RC, Small RL, Haselkorn T, Wendel JF (2002) Rapid diversification of the cotton genus (Gossypium: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. American Journal of Botany 89: 707–725. doi: 10.3732/ajb.89.4.707

Cui XL, Liu F, Liu YL, Zhou ZL, Zhao YY, Wang CY, Wang XX, Cai XY, Wang YH, Meng F, Peng RH, Wang KB (2015) Construction of cytogenetic map of *Gossypium herbaceum* chromosome 1 and its integration with genetic maps. Molecular Cytogeneitcs 8: 2. doi: 10.1186/s13039-015-0106-y

Desel C, Jung C, Cai DG, Kleine M, Schmidt T (2001) High-resolution mapping of YACs and the single-copy gene Hs1pro-1 on Beta vulgaris chromosome by multi-colour fluorescence in situ hybridization. Plant Molecular Biology 45: 113–122. doi: 10.1023/A:1006405911442

Deutsch M, Long M (1999) Intron-exon structure of eukaryotic model organisms. Nucleic Acids Research 27: 3219–3228. doi: 10.1093/nar/27.15.3219

Dong F, Song J, Naess, Helgeson JP, Gebhardt C, Jiang J (2000) Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. Theoretical and Applied Genetics 101: 1001–1007. doi: 10.1007/s001220051573

Ellegren H (2002) Mismatch repair and mutational bias in microsatellite DNA. Trends in Genetics 18: 552. doi: 10.1016/S0168-9525(02)02804-4

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics 9: 397–405. doi: 10.1038/nrg2337

Gan YM, Chen D, Liu F, Wang CY, Li SH, Zhang XD, Wang YH, Peng RH, Wang KB (2011) Individiual chromosome assignment and chromosomal collinearity in *Gossypium thurberi*, *G. trilobum* and D subgenome of *G. barbadense* revealed by BAC-FISH. Genes Genetics System 86: 165–174. doi: 10.1266/ggs.86.165

Gan YM, Liu F, Chen D, Wu Q, Qin Q, Wang CY, Li SH, Zhang XD, Wang YH, Wang KB (2013) Chromosomal locations of 5S and 45S rDNA in *Gossypium* genus and its phylogenetic implications revealed by FISH. PLoS ONE 8: e68207. doi: 10.1371/journal.pone.0068207

Gan YM, Liu F, Peng RH, Wang CY, Li SH, Zhang XD, Wang YH, Wang KB (2012) Individual chromosome identification, chromosomal collinearity and geneticphysical intergrated map in *Gossypium darwinii* and four Dgenome cotton species revealed by BAC-FISH. Genes Genetics System 87: 233–241. doi: 10.1266/ggs.87.233

Gomez MI, Isiam-Faridi MN, Woo SS, Czeschin D Jr, Zwick MS, Stelly DM, Price HJ, Schertz KF, Wing RA (1997) FISH of a maize sh2-selected sorghum BAC to chromosomes of Sorghum bicolor. Genome 40: 475–478. doi: 10.1139/g97-063

Grover CE, Kim HR, Wing RA, Paterson AH, Wendel JF (2004) Incongruent patterns of local and global genome size evolution in cotton. Genome Research 14: 1474–1482. doi: 10.1101/gr.2673204

Grover CE, Kim HR, Wing RA, Paterson AH, Wendel JF (2007) Microcolinearity and genome evolution in the AdhA region of diploid and polyploidy cotton (*Gossypium*). Plant Journal 50: 995–1006. doi: 10.1111/j.1365-313X.2007.03102.x

Hanson RE, Zhao XP, Islam-Faridi MN, et al. (1998) Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). American Journal of Botany 85(10): 1364–1368. doi: 10.2307/2446394

Han YH, Zhang ZH, Huang SW, Jin WW (2011) An integrated molecular cytogenetic map of *Cucumis sativus* L. chromosome 2. BMC Genetics 12: 18–25. doi: 10.1186/1471-2156-12-18

Han ZG, Wang CB, Song XL, Guo WZ, Gou JY, Li CH, Chen XY, Zhang TZ (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. Theoretical and Applied Genetics 112: 430–439. doi: 10.1007/s00122-005-0142-9

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Research 16: 1252–1261. doi: 10.1101/gr.5282906

Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica 115: 49–63. doi: 10.1023/A:1016072014259

Kim JS, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of sorghum bicolor. Genetics 169: 1169–1173. doi: 10.1534/genetics.104.035980

Li FG, Fan GY, Wang KB, Sun FM, Yuan YL, Song GL, et al. (2014) Genome sequence of the cultivated cotton *Gossypium arboretum*. Nature Genetics 46: 567–572. doi: 10.1038/ng.2987

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. Nature Genetics 30: 194–200. doi: 10.1038/ng822

Nguyen TB, Giband M, Brottier P, Risterucci AM, Lacape JM (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theoretical and Applied Genetics 109: 167–175. doi: 10.1007/s00122-004-1612-1

Oliver KR, McComb JA, Greene WK (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biology and Evolution 10: 1886–1901. doi: 10.1093/gbe/evt141

Percival AE, Wendel JF, Stewart JM (1999) Taxonomy and germplasm resources. Wiley Press, New York.

Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O (2006) Doubling genome size without polyploidization: dynamics of retrotransposition driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Research 16: 1262–1269. doi: 10.1101/gr.5290206

Sambrook J, Russell DW (2002) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York.

Senchina DS, Alvarez I, Cronn RC, Liu B, Rong JK, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Molecular Biology and Evolution 20: 633–643. doi: 10.1093/molbev/msg065

Shahmuradov IA, Akbarova YY, Solovyev VV, Aliyev JA (2003) Abundance of plastid DNA insertions in nuclear genomes of rice and Arabidopsis. Plant Molecular Biology 52: 923–934. doi: 10.1023/A:1025472709537

Sun JY, Zhang ZH, Zong X, Huang SW, Li ZY, Han YH (2013) A high-resolution cucumber cytogenetic map integrated with the genome assembly. BMC Genomics 14: 461–468. doi: 10.1186/1471-2164-14-461

Tang X, De Boer JM, Van Eck HJ, Bachem C, Visser RG, De Jong H (2009) Assignment of genetic linkage maps to diploid Solanum tuberosum pachytene chromosomes by BAC-FISH technology. Chromosome Research 17: 899–915. doi: 10.1007/s10577-009-9077-3

Vinogradov AE (1999) Intron-genome size relationship on a large evolutionary scale. Journal of Molecular Evolution 49: 376–384. doi: 10.1007/PL00006561

Wang CY, Wang KB, Song GL, Li MX, Bie S, Li SH, Zhang XD (2001b) Protocol of cotton FISH of somatic chromosomes with rDNA as probes. Cotton Science 13: 75–77.

Wang K, Guan B, Guo WZ, Zhou BL, Hu Y, Zhu YC, Zhang TZ (2008) Completely distinguishing individual A-genome chromosomes and theirkaryotyping analysis by multiple bacterial artificial chromosome fluorescence hybridization. Genetics 178: 1117–1122. doi: 10.1534/genetics.107.083576

Wang K, Guo WZ, Yang ZJ, Hu Y, Zhang WP, Zhou BL, et al. (2010) Structure and size variations between 12A and 12D homoeologous chromosomes based on high-resolution cytogenetic map in allotetraploid cotton. Chromosoma 119: 255–266. doi: 10.1007/s00412-009-0254-0

Wang K, Guo WZ, Zhang TZ (2007) Development of one set of chromosome-specific microsatellite-containing BACs and their physical mapping in *Gossypium hirsutum* L. Theoretical and Applied Genetics 115: 675–682. doi: 10.1007/s00122-007-0598-x

Wang KB, Wang WK, Wang CY, Song GL, Cui RX, Li SH, Zhang XD (2001a) Studies of FISH and karyotype of *Gossypium barbadense*. Yi Chuan Xue Bao 28: 69–75.

Wang KB, Wang ZW, Li FG, Ye WW, Wang JY, Song GL, et al. (2012) The draft genome of a diploid cotton *Gossypium raimondii*. Nature Genetics 44: 1098–1103. doi: 10.1038/ng.2371

Wendel JF (2000) Genome evolution in polyploids. Plant Molecular Biology 42: 225–249. doi: 10.1023/A:1006392424384

Wu Q, Liu F, Li SH, Song GL, Wang CY, Zhang XD, Wang YH, Stelly D, Wang KB (2013) Uniqueness of the Gossypium mustelinum genome revealed by GISH and 45S rDNA FISH. Journal of Integrative Plant Biology 55: 654–662. doi: 10.1111/jipb.12084

Zhang J (2003) Evolution by gene duplication: an update. Trends in Ecology Evolution 18: 292–298. doi: 10.1016/S0169-5347(03)00033-8

Zhao L, Lv YD, Cai CP, Tong XC, Chen XD, Zhang W, Du H, Guo XH, Guo WZ (2012) Toward allotetraploid cotton genome assembly: integration of a high-density molecular genetic linkage map with DNA sequence information. BMC Genomics 13: 539–555. doi: 10.1186/1471-2164-13-539

Zhao X, Lu JY, Zhang ZH, Hu JJ, Huang SW, Jin WW (2011) Comparison of the distribution of the repetitive DNA sequences in three variants of *Cucumis sativus* reveals their phylogenetic relationships. Genetics and Genomics 38: 39–45. doi: 10.1016/j.jcg.2010.12.005

Zhao XP, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Paterson AH (1998) Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. Genome Research 8(5): 479–492.

Zhao XP, Wing RA, Paterson AH (1995) Cloning and characterization of the majority of repetitive DNA in cotton (*Gossypium* L). Genome 38(6): 1177–1188. doi: 10.1139/g95-156

Zhu L, Smith S, de Lange T, Seldin MF (1999) Chromosomal mapping of the tankyrase gene in human and mouse. Genomics 57: 320–327. doi: 10.1006/geno.1999.5771